Tecnológico de Monterrey

IN 4027: Data Science and Statistical Inference

July, 10th, 2020

**Midterm Exam (Take Home Portion)**

Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Student Number:\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. For the Happiness.csv dataset provided, please compute an OLS model (it is OK to use a Python library for that purpose) to explain the Happiness Score, in terms of the other variables. Compute, just using pandas and numpy, the standard errors at each 0.5 increment of the score. Plot the data, your OLS estimates and the confidence bounds given by the standard errors at each of the 0.5 increments. (40 points)

2. Use MLE to numerically fit the distribution of the Happiness score first to a Shifted Gamma and afterwards to a Lognormal distribution (pdf provided below). Plot PDFs of the estimated distributions over a histogram of the score. (30 points)

Imagen que contiene dibujo

Descripción generada automáticamente

3. Divide the data by region, using one category for Latin America and one category to group all other regions together. Assess statistical significance of the difference in **each** of the Happiness’s score components. Further, perform all computations necessary to justify that your conclusion is **not** dependent on normality of your test statistic. (30 points)